

Urban Concentration and Economic Growth: checking for specific regional effects*

Alain PHOLO BALA[†]

May 2009

Abstract

By using a semiparametric specification, we examine the impact of urban concentration in economic growth on different groups of countries that we classify according to a geographical criterion or according to their level of development. Facing a significant proportion of missing data, we handle that problem with a multiple imputation algorithm as advised in the statistical literature. Therefore using a Bayesian estimation we obtain parametric coefficients and non parametric curves. Then we may perform Yatchew's tests of equality of non parametric effects to check out if the model specification is the same in the different groups of countries.

Keywords: urban primacy; economic growth; missing data; imputation; semiparametric estimation

JEL Classification: C14; C30; C49; R15

*We thank Luc Bauwens, Dominique Peeters and Jacques Thisse for very detailed comments on a previous version of this paper. We also thank Fausto Galli, Wilfried Koch for helpful comments and discussions. The usual disclaimer applies.

[†]CORE, 34 Voie du Roman Pays 1348 Louvain-La-Neuve, Belgium. Office: (+32)10474353. Fax: (+32)10474301. E-mail: Alain.PholoBala@uclouvain.be.

1 Introduction

Urbanization appears as closely linked to economic development. In any year the simple correlation coefficient between the urbanization rate and the log of per capita GDP is about 0.85 (Henderson, 2003). The intuition behind this stylized fact is clear. As economies develop, relative and absolute changes in demand increase the relative and absolute importance of the industrial and service sectors. These sectors are much less land intensive than the agricultural sector, and they allow easier substitution of non land for land inputs (Moomaw and Shatter, 1996). Therefore, in spite of urban land's high prices, firms of those sectors can cluster in urban areas to take advantage of Marshall's localized economies of scale (Henderson, 1974; Duranton and Puga, 2001; Henderson, 2003).

However, while urbanization is a universal phenomenon triggered by the sectorial shift from agriculture to industry and modern services, its speed seems to vary according to the level of development. The differential rates of growth between urban and rural population in 1950-80 show an inverted U pattern with middle-income countries having the highest rates while those of developed nations and low income countries are the lowest (Mazumdar, 1987). As middle-income countries are, on average, those that benefit from the highest growth rates and face the most drastic changes in their economic structures, this evidence would suggest that urbanization is related to economic development and industrialization (Yuki, 2007).

Yet, since several nations that urbanized the most rapidly, i.e. African nations and Latin American & Caribbean, grew relatively slowly (Mazumdar, 1987; Fay and Opal, 2000) the speed of urbanization seems not to be explained solely by economic growth (Yuki, 2007). This observation holds for urban concentration as well. This is the aspect of urban development that seems to interest economists the most (Henderson, 2003). This is also the one that has triggered the greatest deal of concern and controversy. In 2005, 15 of the 20 urban agglomerations of more than 10 millions inhabitants were located in developing countries. The surge of so many megalopolises in developing countries has been a subject of concern for international policy officials. For a long time international development agencies have suspected megacities of developing countries to be over populated and have considered their alleged 'overconcentration' as detrimental for economic growth.

Such a prejudice may be partly grounded on the analysis made by Todaro and Bairoch. From different analytical frameworks they arrived both at the same conclusion that there is 'excessive' urban concentration in developing countries. Bairoch argues that excessive urban concentration is due to, among other factors, rapid population growth which leads to rural

crowding and stimulates rural to urban migration. In addition, he claims that artificially high urban wages pull a disproportionate part of the population to urban areas. While making a similar claim concerning urban wages, Todaro's analysis differs by its greater emphasis on economically inefficient migration caused by legally and socially determined minimum wage rates and migrants expectations (Moomaw and Shatter, 1996; Todaro, 1969; Bairoch, 1988).

There is however no unanimity regarding that issue. Challenging Todaro and Bairoch's claim, Williamson (1987) asserted that there is no evidence confirming that developing countries are overurbanized, and that urbanization has outpaced industrialization in developing countries. Mera (1973) claimed that the largest metropolitan areas in the world are likely to be less large than the optimum in terms of economic efficiency. Conversely, Ades and Glaeser (1995) found that both population share of the largest city and urbanization outside the main city have negative and significant effects on growth of GDP per capita, reaching the opposite conclusion that 'Large cities generate rent-seeking and instability, not long term economic growth'.

Economic literature firstly formalized the link between urban concentration and economic growth by the Williamson hypothesis. It states that economic development first increases and then decreases spatial concentration within a country, thus exhibiting a bell shaped relationship (Junius, 1999; Williamson, 1965; Alonso, 1980). At early stages of economic development, a country optimizes the use of its physical infrastructure and managerial resources by clustering them in primate and often coastal cities. Such spatial clustering favors information spillovers and knowledge accumulation when the economy is 'information deficient'. Nevertheless, at later stages of development process, deconcentration proceeds for the mere reason that the economy can sustain the spread of economic infrastructure and knowledge resources in the hinterland and because primate cities have become congested areas that are less efficient for economic agents (Henderson, 2003).

The bell shaped relationship has been confirmed by some empirical studies (El-Shakhs, 1972; Alonso, 1980; Wheaton and Shishido, 1981; Junius, 1999; Davis and Henderson, 2003). But it has also been contradicted by others. Richardson and Schwartz (1988) find no support of any link between primacy and economic growth. As Ades and Glaeser (1995), Mutlu (1989) and Moomaw and Shatter (1996) find a negative relationship between urban concentration and economic development. So alternative explanations focusing on non economic factors have been raised. One of them states that cities grow in a parallel way and that spatial concentration is unaffected by urbanization and economic development (Junius, 1999; Black and Henderson, 1999; Eaton and Eckstein, 1997), the distribution of

urban population reflecting simply geography or historic shocks. Another hypothesis, supported by a large strand of the literature, outlines the importance of political institutions and policies in spatial concentration. Ales and Glaeser (1995) even asserts that ‘political forces, even more than economic forces, drive urban centralization’. For political reasons a government may favor one or more cities over others, especially national capitals. Such a favoritism may take several forms: the government may underinvest in interregional transport and telecommunications favoring therefore consumers and producers in the national capital over those in the hinterland (Fujita *et al.*, 1999); it may impose restrictions in the capital and the export/import markets favoring firms located in the capital; finally it may allocate public services preferentially in the national capital.

One reason of the contradictions just outlined may be that countries are likely to show up a great deal of heterogeneity with respect to urbanization and growth patterns. Urbanization qualitative nature appears to vary across countries with on one hand countries experiencing urbanization accompanied by skill upgrading, industrialization, economic growth and the expansion of the urban formal sector and on the other nations experiencing simultaneously an urbanization without modernization, the expansion of the shadow economy and economic stagnation (Yuki, 2007). In the growth empirical literature, the objection has been raised that very different countries are unlikely to be drawn from a common surface as multiple regression assumes and evidence of widespread parameter heterogeneity has been provided (Temple, 1999).

The evidence that a substantially different pattern of urbanization prevails in Sub-Saharan Africa would then raise the interest of the quest for a way of modeling agglomeration economies more suitable for that part of the world. But does such an evidence exist? *Do developing and developed countries diverge with respect to urbanization patterns?* In order to obtain such an evidence, we analyze the relationship between economic growth and urban concentration. Therefore, our analysis is similar to Henderson’s (2003). However, Henderson (2003) assumed that there exists an optimal level of urban concentration, and showed that deviation from that optimum may be very costly. Here, we relax the assumption that the economic growth rate is a concave function of urban concentration. Therefore, as Bertinelli and Strobl (2003), we model economic growth using a semi-parametric function, with the nonparametric term depending on urban concentration. However, contrary to them, we use a Differencing method to perform semi-parametric estimation. This method has the advantage of allowing us to check out subsequently if urbanization patterns differ across group of countries by performing tests of equality of non-parametric functions of different subsamples. Furthermore, our analysis differs from the one of Bertinelli and Strobl

(2003) by the fact that we address explicitly the endogeneity problem.

Another major difference with the earlier literature is the way we handle the data missingness problem. Most of empirical papers in economics handle missing data by listwise deletion i.e. by deleting any observation having at least one missing datum. This approach has faced a lot of criticism by the statistics literature. Indeed, apart from the fact that it implies the loss of all the information conveyed in the observations having missing data, the estimates obtained with such a method have been proved to be biased if the data remaining after deletion is not a random sample of the overall database.¹ In order to avoid such a shortcoming, we implement a two-step multiple imputation algorithm which is convenient if data are missing at random (*MAR*).

The remainder of this paper is organized as follows. Section 2 is devoted to the description of the estimation methodology. In this section we justify our specification choice, we present our basic estimation strategy - the Bayesian Semiparametric estimation - and tests of equality of nonparametric regression functions. Then we discuss endogeneity issues and methods for handling data missingness. Finally we describe our Estimation-Imputation algorithm. Section 3 presents the results obtained and Section 4 concludes.

2 Estimation Methodology

2.1 Specification

The Empirical Growth literature has for some time been dominated by papers with cross-country growth regressions. The formulation and the relevance of such regressions have been quickly subject to a rising skepticism. This approach was indeed prone to several shortcomings. The most important of them is that as cross-sectional regressions fail to control for individual heterogeneity, they face an omitted variables problem and thus yields biased estimates. The use of panel data allows to mitigate such an inconvenience. Indeed, in a panel data framework one may control for heterogeneity in the initial level of efficiency and thus ensure that coefficients will be unbiased. Secondly with panel data several lags of regressors may be used as instruments, alleviating therefore measurement errors and endogeneity biases (Temple, 1999; Magrini, 2004)

However, implementing traditional growth regression in a panel framework has some major drawbacks : rather than having exogenous technological change and population growth, they include determinants of population change leading away from the standard

¹If the converse assertion were true data would be described as missing completely at random (*MCAR*).

neo-classical framework. Moreover, estimation of growth models implies the additional complexities of dynamic panel data models (Henderson, 2003; Temple, 1999). So we estimate a Total Factor Productivity model. Total Factor Productivity models rely on a production function with two factors of production, physical capital K_{it} and labor L_{it} , and a variable reflecting the effects of technological progress A_{it} . Adopting the Cobb-Douglas specification, we get the following production function:

$$Y_i(t) = K_i(t)^\alpha A_i(t)(L_i(t))^{1-\alpha} \quad (1)$$

Linearizing, differencing and normalizing with respect to labor leads to the following specification:

$$\ln \frac{Y_{i(t)}}{L_{i(t)}} - \ln \frac{Y_{i(t-1)}}{L_{i(t-1)}} = \alpha [\ln \frac{K_{i(t)}}{L_{i(t)}} - \ln \frac{K_{i(t-1)}}{L_{i(t-1)}}] + \ln \frac{A_{i(t)}}{A_{i(t-1)}} \quad (2)$$

Total factor productivity growth is modeled as a function of (i) education of the labor force, which captures the capacity of adopting new technologies (Grossman and Helpman, 1991; Durlauf and Quah, 1998), (ii) internal country considerations affecting efficiency and growth, like urban primacy (Henderson, 2003). We therefore assume the following functional form for productivity growth

$$\ln(A_{i(t)}/A_{i(t-1)}) = z'_{i(t-1)}\delta + f(x_{i(t-1)}) \quad (3)$$

with x representing urban concentration, and z representing a row vector of control variables including: average years of high school and college as a proxy of education of the labor and time fixed effects.²

We consider the second term of the right hand side of equation (3) as a non-linear function of urban concentration. There are indeed good reasons to assume that it is not linear in urban concentration. Firstly there is the so-called Williamson effect according to which urban concentration should be high at first stages of development and then decrease as the economy develops. The Williamson effect implies therefore a bell shaped relationship between urban concentration and the level of economic development. Henderson (2003) has checked the validity of Williamson hypothesis by modeling $f(x)$ as a quadratic function

²In this semi-parametric specification we do not control for countries fixed effects. Indeed, the differencing method we use for performing this estimation implies a reordering of the data matrix that undermines the implementation of efficient fixed effects estimation methods. Therefore, to estimate fixed effects we must add countries dummies to the matrix of regressors, which increases remarkably its size and the number of coefficients to estimate. Thus, we only control for fixed effects when we estimate the model without the non-parametric function as shown in table 2. This omission of fixed effects is likely to trigger endogeneity. We will present a solution to this problem when we will address endogeneity issues.

of x . With such a function he was also able to verify that, as we may expect, urban concentration decreases with the population of a country and national geographic size and he has shown that there is an optimal urban concentration level from which departures could entail significant losses in terms of economic growth. Secondly as noticed by Bertinelli and Strobl (2003), urban concentration variables are bounded from above and below, and may even just affect growth differently near their bounds comparatively to mid-values.

While the intuition behind the Williamson effect is quite appealing, as stated previously there are compelling arguments suggesting other kinds of relationship. So we may wonder if the Williamson hypothesis is actually supported by the data. Thus, following Bertinelli and Strobl (2003) we will allow some flexibility in our specification by using a semiparametric specification, with the nonparametric term being a function of urban concentration.

2.2 Bayesian Semiparametric estimation

Because of the plurality of theoretical frameworks as well as diverging empirical results, we have no obvious functional form for the regression relationship between growth rates, urban concentration and other controls. Such a context gives backing to arguments of non-parametric econometricians who stress that the implications of economic theory is often non-parametric and propose semiparametric or non-parametric functional forms (Koop and Poirier, 2004; Yatchew, 1998). Therefore, we opt for working with the following semiparametric specification:

$$\Delta \ln(Y_{i(t)}/L_{i(t)}) = z'_{1i(t)}\beta + f(x_{i(t-1)}) + \epsilon_{i(t)}, \quad \epsilon_{i(t)} \sim \mathcal{N}(0, \sigma^2) \quad (4)$$

Where $z_{1i(t)} = (\Delta \ln(K_{i(t)}/L_{i(t)}), z_{i(t-1)})'$, and $\beta = (\alpha, \delta)'$.

However, nonparametric methods are not very popular in applied work for the mere reason that nonparametric regression techniques are theoretically more complex than the usual tool kit of linear and nonlinear parametric modeling methods, and that they are computationally intensive (Yatchew, 1998).

To avoid such shortcomings, we use a Bayesian Semiparametric model (Koop and Poirier, 2004). Based on the standard Normal linear regression model with natural conjugate prior for which standard analytical results are available, this model has the advantage to be computationally simple. Furthermore, by using such a method we avoid the criticism addressed to usual Bayesian methods that they incorporate prior information. Indeed, in the approach we are using, the only type of prior input required is one prior hyperparameter, η , which controls the degree of smoothness of $f(x)$.

Let's consider N as the number of cross-sectional units and T as the number of time periods for each of them. In Bayesian semiparametric estimation of (4) $f(x)$ plays the role of an intercept and observations have to be reordered so that $x_1 \leq \dots \leq x_i \leq \dots \leq x_{N(T-1)}$. Defining $y = (\Delta \ln(Y_1/L_1), \dots, \Delta \ln(Y_{N(T-1)}/L_{N(T-1)}))'$, $x = (x_1, x_2, \dots, x_{N(T-1)})'$, $Z = (z_{11}, \dots, z_{1N(T-1)})'$, and letting $\gamma = (f(x_1), \dots, f(x_{N(T-1)}))'$, $W = (Z, I_{N(T-1)})$ and $\theta = (\beta', \gamma')$, we can rewrite (4) as:

$$y = W\theta + \epsilon, \quad (5)$$

In equation (5) there are more variables than observations. Therefore, additional information is needed to overcome the fact that $W'W$ is singular. Assuming that $f(x_{i(t-1)})$ is smooth we may use the following partially informative prior (Koop and Poirier, 2004):

$$D\gamma \sim \mathcal{N}_{N(T-1)-m}(0, \sigma^2 P_0^{-1}) \quad (6)$$

where $P_0^{-1} = \text{diag}(\eta \ \eta \ \dots \ \eta)$ and D is a differencing matrix given by the following expression:

$$D_{(N(T-1)-m) \times N(T-1)} = \begin{bmatrix} d_0, d_1, d_2, \dots, d_m, 0, \dots, 0 \\ 0, d_0, d_1, d_2, \dots, d_m, 0, \dots, 0 \\ \vdots \vdots \\ 0, \dots, 0, d_0, d_1, d_2, \dots, d_m, 0 \\ 0, \dots, 0, d_0, d_1, d_2, \dots, d_m \\ 0, \dots, 0 \\ \vdots \vdots \\ 0, \dots, 0 \end{bmatrix} \quad (7)$$

with m the order of differencing, d_i with $0 \leq i \leq m$ a differencing weight. Combined with an Inverted Gamma prior on σ^2 , the prior on $D\gamma$ is conjugate. It yields an integrable posterior. Indeed, the posterior precision $W'W + M_0$ is of full rank.³ To estimate β we apply the differencing matrix to (5) which yields:

$$Dy = DZ\beta + Df(x) + D\epsilon \approx DZ\beta + D\epsilon \quad (8)$$

Application of OLS to (8) is straightforward.

Once $\hat{\beta}$ is obtained, we may obtain γ by the following extended regression:

$$\begin{pmatrix} y - Z\hat{\beta} \\ 0_{N(T-1)-m} \end{pmatrix} = \begin{pmatrix} I_{N(T-1)} \\ (P_0^{1/2})' D \end{pmatrix} \gamma + \begin{pmatrix} \epsilon \\ \epsilon_0 \end{pmatrix} \quad \text{with} \quad \begin{pmatrix} \epsilon \\ \epsilon_0 \end{pmatrix} \sim \mathcal{N}(0, \sigma^2 I_{2N(T-1)-m}) \quad (9)$$

³the prior precision M_0 being given by $M_0 = \begin{pmatrix} 0 & 0 \\ 0 & P_0 \end{pmatrix}$

2.3 Tests of equality of regression functions

With such a nonparametric estimation framework we may perform tests of equality of nonparametric functions across subsamples. Several procedures to realize such tests have been designed. Most of them involve direct comparison of nonparametric estimates of regression curves or analysis of residuals from such regressions. Yatchew (1999) eases the test by proposing a procedure avoiding the computation of nonparametric regressions. Suppose that we have G subsamples of different sizes respectively N_1, N_2, \dots, N_G with $\sum_{i=1}^G N_i = N(T-1)$. Let's apply regression model (5) to each subsample separately, then we have for any subsample i we have:

$$y_i = Z_i\beta + f(x_i) + \epsilon_i, \quad (10)$$

where $y_i = (y_{i1} \dots y_{iN_i})'$, $Z_i = (z_{i1}, \dots, z_{iN_i})'$, $x_i = (x_{i1}, \dots, x_{iN_i})'$, $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iN_i})'$ with $\epsilon_i \sim \mathcal{N}(0, \sigma^2 I_{N_i})$ and $i = 1, \dots, G$.

Supposing that data have been reordered so that within each subsample, the x 's are in increasing order and defining $y = (y'_1, \dots, y'_G)'$, $Z = (Z'_1, \dots, Z'_G)'$ and $y_{npar} = y - Z\beta$ then we may compute the 'within' estimator of σ^2 as:

$$s_w^2 = \frac{1}{N(T-1)} y'_{npar} D'_{test} D_{test} y_{npar}, \quad (11)$$

with D_{test} defined to be the following block diagonal matrix

$$D_{test} = \begin{bmatrix} D_1 & 0 & \dots & 0 \\ 0 & D_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & D_G \end{bmatrix} \quad (12)$$

with each bloc D_i of dimension $N_i \times N_i$ having the same structure as (7).

Let's define P_p as the 'pooled' permutation matrix that reorders the overall dataset so that the 'pooled' x are in increasing order. Then we define the 'pooled' variance estimator as:

$$s_P^2 = \frac{1}{N(T-1)} y'_{npar} P'_p D'_{test} D_{test} P_p y_{npar}, \quad (13)$$

Under the null hypothesis that all nonparametric regression functions are identical we may define the following test statistic

$$\Upsilon = (mN(T-1))^{1/2} (s_P^2 - s_w^2) = \frac{m^{1/2}}{(N(T-1))^{1/2}} y'_{npar} Q_\Upsilon y_{npar} \xrightarrow{D} \mathcal{N}(0, 2\pi_\Upsilon \sigma^4), \quad (14)$$

with $Q_{\Upsilon} = P_p' D_{test}' D_{test} P_p - D_{test}' D_{test}$ and $\hat{\pi}_{\Upsilon} = m \text{tr}(Q_{\Upsilon} Q_{\Upsilon}) / NT$ and supposing that $\hat{\pi}_{\Upsilon} \xrightarrow{P} \pi_{\Upsilon} > 0$.

Therefore, $\Upsilon / s_w^2 (2 \hat{\pi}_{\Upsilon})^{1/2} \xrightarrow{D} \mathcal{N}(0, 1)$ and we would reject the null hypothesis for large positive values of the test statistic. Rejection of the null hypothesis suggests that the pattern of urbanization is different between groups of countries.

2.4 Data

Several variables are involved in estimation of equation (5) namely education of the labor force, urban concentration, capital and output growth rates. While measuring education is quite straightforward with average years of high school and college of population of at least 25 years old standing as a convincing proxy, things are more involved with respect to urban concentration. There have always been several measures of spatial concentration so that deciding what measure to opt for is an issue. The Hirschman-Herfindalh index and the Pareto parameter have been the first to be used. But there are available only for few years for a limited sample of countries. Therefore, they don't fit for panel data.

Urban primacy defined as the share of urban population living in the largest city, is conversely available over years for more countries. It is moreover closely correlated to the previous measures (Henderson, 2003) and thus has been used in many studies. But as explained in Bertinelli and Strobl (2003), this measure seems unsatisfactory when there are huge differences between country sizes. Indeed, small countries tend to gather the quasi-totality of their urban population in a single city. Moreover, there are often cities other than the largest city that account for large proportions of the urban population. Therefore, using urban primacy as a measure of urban concentration results in attributing low values to countries like India which has many large cities, but very large values for small countries. Furthermore, changes in urban primacy sometimes don't reflect changes in the total population. It has been noticed that while the share of the largest city often decreased as a consequence of increasing urbanization, urban concentration increased due to a more than proportional increase in medium and large agglomerations. All those shortcomings have induced Bertinelli and Strobl (2003) to adopt another measure of urban concentration: urban density defined as the share of the urban population living in cities larger than 750,000 inhabitants. For reason of completeness we will adopt both measures in this study. This will allow us to check whether the results are robust to changes in measures.

Estimation of a TFP model requires data on the capital stock. We use Dareshwar and Nehru (1993) data on the capital per capita along with their output per worker measure.

Those measures are based on perpetual inventory methods and are in local currency units. To take into account variations of purchasing power across countries, those results were converted in PPP at 1987 exchange rate.

2.5 Endogeneity Issues

The use of Dareshwar and Nehru data raises endogeneity issues. Indeed, although their measurements are carefully done, there are likely to suffer from measurements errors.

Endogeneity issues are further compounded by the fact that in equation (4) contemporaneous shocks $\epsilon_{i(t)}$ potentially affect covariates at period t and even at period $t - 1$. Indeed, perspectives of shocks in economic growth are likely to induce migration to the largest city, increasing therefore urban concentration (Henderson, 2003). Finally this endogeneity problem is also triggered by the fact that equation (4) implies a pooled estimation where individual heterogeneity is not controlled for. Since fixed effects are likely to be correlated to regressors this fuels the endogeneity bias.

Therefore, regressors are not strictly exogenous. In order to be able to identify the ‘causal’ effect of urban concentration on economic growth rather than simple correlations we have to explicitly address this endogeneity problem. To do so using values of covariates at $t - 2$ and $t - 3$ as instruments, we may implement instrumental variables techniques. Such a task may appear more involved in the context of semiparametric estimation since conventional instrumental variables techniques seem to be not directly transferable in a semiparametric framework. However, Yatchew (2003) presents an approach for handling endogeneity in nonparametric estimation. Using that approach, let’s denote w as a vector of instruments for $x_{(t-1)}$ uncorrelated with $\epsilon_{i(t)}$, with w being given by $w_{(t-1)} = (x_{(t-2)}, x_{(t-3)})'$, then:

$$x_{(t-1)} = w'_{(t-1)}\pi + u_{(t-1)} \quad E(u_{(t-1)}|w_{(t-1)}) = 0 \quad E(\epsilon_{(t)}|w_{(t-1)}) = 0 \quad (15)$$

Suppose now that $E(\epsilon_{(t)}|x_{(t-1)}) = \rho u_{(t-1)}$ so that $\epsilon_{(t)} = \rho u_{(t-1)} + v_{(t)}$. We can thus rewrite equation (4) as:

$$\Delta \ln(Y_{i(t)}/L_{i(t)}) = z'_{2i(t)}\beta_1 + f(x_{i(t-1)}) + v_{i(t)}, \quad v_{i(t)} \sim \mathcal{N}(0, \sigma^2) \quad (16)$$

with $z_{2i(t)} = (z'_{1i(t)}, u_{(t-1)})'$, $\beta_1 = (\beta', \rho)'$, and $E(v_{(t)}|x_{(t-1)}, z_{2(t-1)}) = 0$.

After differencing equation (16), we may apply instrumental variables estimation to the parametric variables only, moving therefore from an endogeneity problem in a nonparametric estimation to an endogeneity problem in a parametric specification.

Rewriting (16) as

$$Dy = DZ_2\beta_1 + Df(x) + D\epsilon \approx DZ_2\beta + D\epsilon \quad (17)$$

where $y = (\Delta \ln(Y_1/L_1), \dots, \Delta \ln(Y_{N(T-1)}/L_{N(T-1)}))'$, $Z_2 = (z_{21}, \dots, z_{2N(T-1)})'$, $x = (x_1, x_2, \dots, x_{N(T-1)})'$,

we may rearrange (17) in order to distinguish regressors that are endogenous and those that are predetermined. Thus, we get:

$$Dy = DY_1\beta_2 + DZ_{11}\gamma + Df(x) + D\epsilon \approx DY_1\beta_2 + DZ_{11}\gamma + D\epsilon \quad (18)$$

where Y_1 is a $N(T-3) \times m_1$ data matrix gathering all endogenous regressors: stock of capital per capita, average years of schooling and u_{t-1} , the residual of the regression of x_{t-1} , on its instruments w_{t-1} , and Z_{11} is a $N(T-3) \times k_1$ data matrix collecting all predetermined regressors i.e. time dummies.

Equation (18) is only one of the structural equations of a static *simultaneous equations model* (SEM) containing as much equations as endogenous variables. We may estimate parameters of (18) in a ‘limited information spirit’, i.e. without explicit consideration to the restrictions pertaining to the remaining structural equations. To do so it is necessary to join to (18) the reduced form corresponding to endogenous variables that appear as regressors in (18)

$$DY_1 = DZ_1\Pi + DV_1 = DZ_{10}\Pi_0 + DZ_{11}\Pi_1 + DV_1 \quad (19)$$

where $Z_1 = (Z'_{10}, Z'_{11})'$ stands for the $N(T-3) \times k$ matrix gathering predetermined regressors as well as instruments of endogenous variables, gathered in data matrix Z_{10} .

Thus, we have to estimate the following system of equations

$$Dy = DY_1\beta_2 + DZ_{11}\gamma + D\epsilon \quad (20)$$

$$DY_1 = DZ_{10}\Pi_0 + DZ_{11}\Pi_1 + DV_1 \quad (21)$$

2.6 Method for handling data missingness

The dataset we are working on is characterized by a significant rate of data missingness. From the 679 observations corresponding to 97 countries, 204, i.e. 30% observations have

at least one missing value. Moreover, 3 of the 5 variables have missing observations.⁴

Table 1: List of variables with missing values

N°	Variable	Missing values	% Missing
1	economic growth rate	158	23
2	capital growth rate	158	23
3	total average year of schooling	117	17

The general way to handle missing data is to transform the incomplete dataset into a complete one. The usual practice to artificially create a complete data set implies either: throwing away cases with missing values (listwise deletion), or imputing, i.e., estimating and filling in, missing data using some ad hoc method like mean imputation, regression-based imputation, dummy variable adjustment, hot-decking.... Then one treats the altered data set as if the deleted cases had never been observed, or the imputed values had always been observed (Schimert *et al.*, 2001).

Listwise deletion and ad hoc methods can lead to misleading inferences because they either throw away or distort information in the data. Listwise deletion is the usual practice for handling data missingness in empirical research in economics. Yet, by throwing away information, listwise deletion may cause at best a significant loss of information, and they may even induce a severe selection bias if data are not *Missing Completely at Random* (MCAR), i.e. missing data are not a random sample of the complete dataset (Schafer, 1997; King *et al.*, 2001). Ad hoc imputation methods don't fix the problem. For instance imputing averages on a variable-by-variable basis preserves the observed sample means, but it distorts the covariance structure. On the other hand imputing predicted values from regression models inflates observed correlations. More generally ad hoc imputations methods by treating imputed data as if there were real fail to reflect any uncertainty due to missing data and thus produce biased standard errors, and p-values (Schafer, 1997).

Therefore, the appropriate way to handle missing data is then to rely on model-based imputation methods. In that category multiple imputation has a clear advantage over single imputation. Indeed, conversely to the latter, it yields inference reflecting sampling variability due to the missing values (Schimert *et al.*, 2001). Multiple imputation methods are unbiased if the missing-data mechanism is ignorable.⁵

⁴Time dummies are not included.

⁵A missing-data mechanism is ignorable if it is *Missing at Random* (MAR), i.e the probability that

Schafer (1997) compares model based imputation and listwise deletion by evaluating the performance of Maximum Likelihood (ML) estimates and listwise deletion (CC for Complete Case) estimates by simulation. To do so he consider a bivariate dataset with variable Y_1 completely observed for units $1, 2, \dots, n$ and Y_2 observed only for units $1, 2, \dots, n_1$ with $n_1 < n$ the number of observations.

He showed that the CC estimate is biased whenever $\rho \neq 0$ for the non-MCAR missingness mechanisms whereas the ML estimates are unbiased under all the missingness mechanisms.⁶ Moreover, under the more restrictive MCAR, ML estimates have an advantage over CC estimates whenever $\rho \neq 0$ because their variances are lower. The explanation for this low variance is that Y_1 becomes an increasingly valuable predictor of the missing values of Y_2 as ρ increases. Therefore, from considerations of bias, consistency and efficiency ML estimates are superior to CC estimates.

Model-based multiple imputation methods assume a statistical model for the distribution of data. The model that is the most widely used is the joint multivariate normal model. It requires special iterative computation tools to extract meaningful summaries like parameters estimates and standard errors. Those computation tools proceed generally in two steps. First, conditionally on the observed values and a starting value of θ - the parameters matrix -, missing values are imputed. Then, once missing values are imputed, a complete dataset is obtained from which θ can be computed. Given this value of θ we can reperform the first step and so on until the algorithm converges (Schafer, 1997). Two principal classes of algorithm are generally considered in the context of model-based multiple imputation methods: Data Augmentation (DA) and Expectation-Maximization (EM) algorithm and extensions.

In this paper we implement a Data Augmentation algorithm. But we can't use it as such. In fact variables in the regression model (5) differ merely by the fact that some of them are independent variables and one is a dependent variable. Depending on the kind of variable we are dealing with, different distributions should be considered. While the standard multivariate model may hold for the group of independent variables, it is more logic to impute the dependent variable according to distribution implied by the regression

datum is missing may depend on the datum itself but only through variables that are observed and if the parameters $\theta = \text{vec}(\mu, \Sigma)$ of the data model and the parameters ϕ of the missingness mechanism are distinct. Indeed, under ignorability, neither the model of the missingness mechanism nor the nuisance parameters ϕ are relevant for making inferences about θ (Schafer, 1997).

⁶ ρ is the correlation coefficient.

model (5), i.e considering that:

$$y \sim \mathcal{N}(W\theta, \sigma^2 I_{N(T-1)}) \quad (22)$$

But then the imputation process becomes much more involved since the dependent variables should be imputed conditionally on the independent variables that have also to be imputed. To perform such a two-step imputation we implement a Gibbs-Sampler algorithm.

2.7 Estimation-Imputation algorithm

As indicated above our algorithm is divided in two steps: firstly imputation of missing covariates and secondly estimation of parameters and imputation of the dependent variable.

2.7.1 Imputation of covariates by a data augmentation algorithm

The Data augmentation algorithm is merely a Gibbs sampler implying two steps that are performed iteratively: an I-step and a P-step.⁷ The I-step simulates missing values of covariates, given observed values of covariates and values of parameters of the data matrix computed at the preceding iteration

$$X_{mis}^{(t+1)} \sim P(X_{mis} | X_{obs}, \theta^{(t)}), \quad (23)$$

while the P-step simulates values of the data matrix parameters at the current iteration given observed values of covariates and missing values of covariates computed at the current iteration

$$\theta^{(t+1)} \sim P(\theta | X_{obs}, X_{mis}^{(t+1)}). \quad (24)$$

2.7.2 Bayesian estimation and imputation of missing values of the dependent variable

Without any endogeneity concern, imputation of the dependent variable is once again a Gibbs Sampler involving two steps: firstly imputation of the missing values of the dependent variable according to the regression model and conditionally on observed values of the regressand, on values of the regressors and on parameters of the regression model computed at the previous iteration:

$$y_{mis}^{t+1} | y_{obs}, X, \beta^t, \sigma_t^2 \sim \mathcal{N}(X\beta^t, \sigma_t^2) \quad (25)$$

⁷A detailed description of this algorithm is provided in Appendix C. The algorithm pseudocode is described in Appendix E.

and secondly draws of the parameters from a Normal-Inverted Gamma distribution conditionally on observed values of the dependent variable, on values of the dependent variable imputed at the current step, and on the regressors

$$\beta^{t+1}, \sigma_{t+1}^2 | y_{obs}, X, y_{mis} \sim NIG(\hat{\beta}, X'X, s, n - k - 2) \quad (26)$$

which implies firstly drawing σ^2 from an Inverted Gamma distribution and then conditionally on σ^2 drawing β from a Normal distribution

$$\beta^{t+1} | \sigma_{t+1}^2 \sim \mathcal{N}(\hat{\beta}, \sigma_{t+1}^2 (X'X)^{-1}) \quad (27)$$

$$\sigma_{t+1}^2 \sim IG_2(n - k - 2, s) \quad (28)$$

With endogenous regressors things are more involved. There is a huge amount of literature on Bayesian limited information estimation of SEM (Dreze and Richard, 1983; Zellner *et al.*, 1988; Bauwens and Van Dijk, 1990). Dreze and Richard (1983) provided results on exact Bayesian analysis of SEM and showed that for a specific choice of prior the posterior distribution of parameters is a poly-t density. The problem is that such a distribution is generally not analytically tractable. It must be integrated numerically to obtain moments, marginal distribution, etc. Moreover, it does not have simple forms from which draws of structural parameters can be made easily.

Zellner *et al.* (1988) provide an alternative approach to deal with Bayesian limited information estimation which avoids such shortcomings by allowing direct Monte Carlo simulation. Following their approach we estimate structural coefficients of equation (18) by a Gibbs Sampler. In the first step this algorithm draws the vector of structural parameters from a multivariate student conditionally on data and on the reduced form parameters. Then on a second step it draws reduced form parameters conditionally on data and on structural form coefficients. Let's rewrite (20) as

$$Dy = W_2 \delta_2 + D\epsilon. \quad (29)$$

The first step draws the vector of structural form parameters δ_2 from $p(\delta_2 | \Pi, D)$ which is a multivariate student density having as parameters

$$\tilde{\delta}_2 = (W_2' M_{V_1} W_2)^{-1} W_2' M_{V_1} Dy \quad (30)$$

$$s_1^2 = \frac{1}{\nu_1} (Dy - W_2 \tilde{\delta}_2)' M_{V_1} (Dy - W_2 \tilde{\delta}_2) \quad (31)$$

with $\nu_1 = N(T - 3) - (m_1 + k_1)$ the number of degrees of freedom and

$$M_{V_1} = I_{N(T-3) \times N(T-3)} - DV_1(V_1' D' DV_1)^{-1} V_1' D'$$

The second step draws the matrix of reduced form parameters Π from $p(\Pi|\delta_2, D)$ which is a matricvariate student density having as formula

$$p(\Pi|\tilde{\delta}_2, D) = f_{MT}^{k \times m_1}(\Pi|\tilde{\Pi}, Y_1' \tilde{M}_\epsilon Y_1, Z' M_\epsilon Z, N(T-3) - k), \quad (32)$$

where $\tilde{\Pi} = (Z' M_\epsilon Z)^{-1} Z' M_\epsilon Y_1$ and

$$\tilde{M}_\epsilon = M_\epsilon - M_\epsilon Z (Z' M_\epsilon Z)^{-1} Z' M_\epsilon, \quad M_\epsilon = I_{N(T-3) \times N(T-3)} - D\epsilon(\epsilon' D' D\epsilon)^{-1} \epsilon' D'$$

We estimate parameters at each step by the means of each conditional distributions considered in the previous approach rather than drawing from those distributions. This procedure corresponds merely to performing 2SLS estimation at each iteration.

3 Results

3.1 Convergence of the missing data imputation algorithm

To ensure convergence of the overall algorithm we ran 2000 iterations.⁸ Several convergence checks were necessary to select a number of iterations allowing us to obtain reliable imputations of regressors missing data. Figures 8 and 9 in Appendix D depict results for various diagnosis of convergence for respectively 2000 and 5000 iterations. As we see even for 2000 iterations autocorrelation between draws dies out very quickly, the graphs of the standardized CUMSUM statistics (convergence graphs) converge smoothly to zero, and plots of draws do not show any long run tendency. Diagnosis of convergence for 5000 iterations confirm that convergence is definitely achieved after 2000 iterations since no excursions away from zero are observed after that number of draws. Therefore, we opt to fix the number of iterations of the imputation algorithm to 2000.

3.2 Estimation results

3.2.1 Basic productivity model

We start our presentation of the results by showing the outcomes of the estimation of a basic productivity model excluding any urban concentration variable. In such a model the only argument in the productivity growth function $\ln(A_{i(t)}/A_{i(t-1)})$ is education of the labor force. Table 2 presents those baseline results.

⁸In Appendix D we describe the criteria used to assess convergence.

Table 2: Estimation of a basic productivity growth equation. dependent variable is $\Delta \ln(Y_{(t)}/L_{(t)})$.

	(1) Fixed effects	(2) IV
$\Delta \ln(K_{(t)}/L_{(t)})$	0.5034**	0.2393**
Human capital	0.0006	0.0183**
Years effects	Yes	Yes
N[countries]	582[97]	388[97]

The multiple imputation method implemented allows a higher data coverage than the one obtained with listwise deletion methods. As only 68 countries have complete information, casewise deletion methods would perform estimation only on a much lower number of observations. Henderson (2003) estimates that basic productivity model from only 482 observations corresponding to 82 countries. Results of Fixed Effects estimation procedures yield estimates of elasticity of capital that are much higher than capital coefficient reported by the literature. Furthermore, coefficient of the human capital proxy is non significant. While the literature indicates that results on the education variable are non robust (Temple, 1999), those obtained on capital elasticity are clearly questionable. Hall and Jones (1999) assumes 0.33 as a capital coefficient and estimates provided by Henderson *et al.* (2001) from work on Korea lie in the range of 0.37-0.39. IV method obtains different results. With an estimate of elasticity of capital of 0.239, IV result is even lower than what is generally accepted by the literature. Moreover, the estimate of human capital coefficient (0.0183) has become significant.

By looking at table 3, we can notice the gap between estimates obtained with the listwise deletion method and earlier estimates obtained with imputation. For OLS estimation, listwise deletion implies higher estimates for both independant variables. For instrumental variables estimation, it entails a higher estimate of the elasticity of capital and a lower estimate of the human capital regression coefficient. Moreover, table 3 clearly indicates the significant loss of degrees of freedom resulting from the use of casewise deletion method: only 69 countries provide relevant observations for estimation. This outlines the payoff yielded by the use of imputation methods.

Table 3: Estimation of a basic productivity growth equation with the listwise deletion method. Dependent variable is $\Delta \ln(Y_{(t)}/L_{(t)})$.

	(1) Fixed effects	(2) IV
$\Delta \ln(K_{(t)}/L_{(t)})$	0.5163**	0.2729**
Human capital	0.0035	0.0161**
Years effects	Yes	Yes
N[countries]	406[69]	268[69]

Table 4: Estimation of a productivity growth equation with primacy. Dependent variable is $\Delta \ln(Y_{(t)}/L_{(t)})$.

	(1) OLS	(2) IV
$\Delta \ln(K_{(t)}/L_{(t)})$	0.5221**	0.3402**
Human capital	0.0169**	0.0169**
Years effects	Yes	Yes
N[countries]	582[97]	388[97]

3.2.2 Productivity model with primacy

Parametric estimates

Table 4 presents results of a productivity model including primacy as the urban concentration measure. As before OLS estimation procedure provides a capital coefficient that is much higher than commonly found in the literature. Conversely, to the previous estimation the human capital coefficient is now significant. Yet, its value is much weaker than the one obtained by Henderson' (2003) GMM estimation.

Instrumental variables estimation yields more appealing results for the parametric coefficients. The elasticity of capital is now in line with the literature. Our results are even closer to literature than estimates provided by Henderson *et al.* (2001) from work on Korea which lie in the range of 0.37-0.39 and are thus higher. Since OLS and IV procedures yield equivalent values of the human capital coefficient, the former does not seem to suffer from a significant endogeneity bias.

Non-parametric curves

For OLS procedure, estimation of the non-parametric regression curve for the overall sample provides a so irregular pattern that no clear lessons can be drawn from it. Conversely, when we consider sub-samples constituted from geographical or developmental criteria, procedures provide interesting results. Sub-samples non-parametric results are reliable only with imputation. Indeed, as table 12 shows, regions such as Asia and Sub-Saharan Africa have a significant share of observations with missing data (respectively 32.92% and 48.45%). The resulting reduction of sample sizes would preclude meaningful non-parametric estimation.⁹

The pattern shown by the non-parametric curve estimated for Europe (Figure 1) is quite irregular. However, except for values of primacy below 0.2 where GDP decreases with primacy or lying in the range 0.23-0.39 where the curve is flat, the curve exhibits a positive slope. This slope is even the sharpest for the 0.20-0.23 and the 0.39-0.46 ranges. This curve suggests that except for low values of primacy, GDP is globally an increasing function of primacy. So while low values of primacy appear to be detrimental to economic growth, urban concentration seems to be associated to economic growth for most of its variation range. A result that recalls findings of Wheaton and Shishido (1981) with a sample gathering countries from all regions.

Countries from other continents exhibit a very different picture as shown by non-parametric curves drawn from other regions of the World. Figure 1 also shows the non-parametric curve obtained from Asia. This curve is irregular and exhibits a visible urban concentration trap.

Latin America non-parametric curve appears as the most irregular, indicating that it is difficult to draw out a smooth non-parametric regression function from this subsample (Figure 2). This non-parametric curve declines sharply after a level of primacy higher than 0.30. Latin America has been for a long time the region where primacy concerns are the highest. This result indicates that the huge urban concentrations prevailing in that region are detrimental to economic growth confirming the prejudice that most of countries of that region are overconcentrated.

Africa provides an interesting picture. It is the only continent that exhibits the bell shaped predicted by Williamson hypothesis. Economic growth increases sharply with primacy until a maximum of about 0.38 is reached then it declines till a primacy level equal to

⁹Table 12 shows that Europe has a significant share of missing data as well (28.57 %). This is caused by the poor information obtained from countries of Eastern Europe for the GDP per capita and the capital per worker variables.

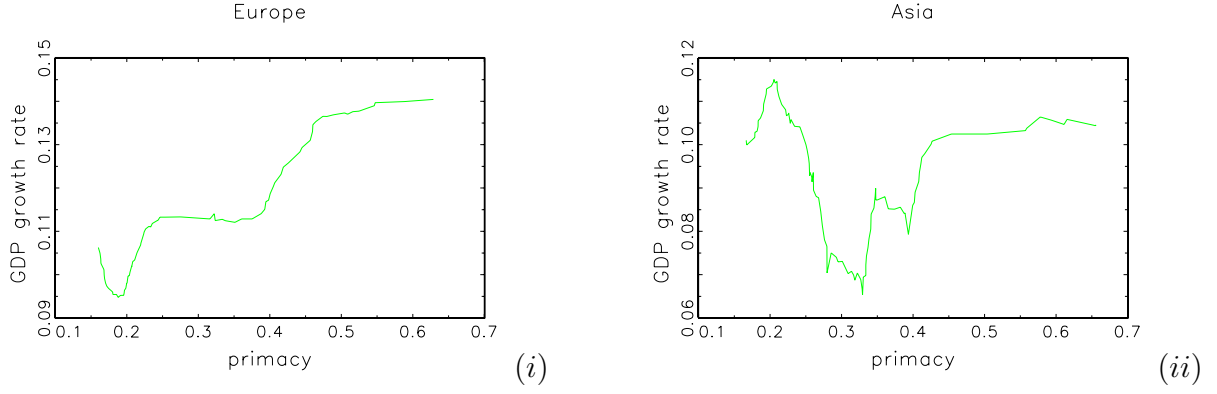


Figure 1: European and Asian non-parametric curves. OLS procedure

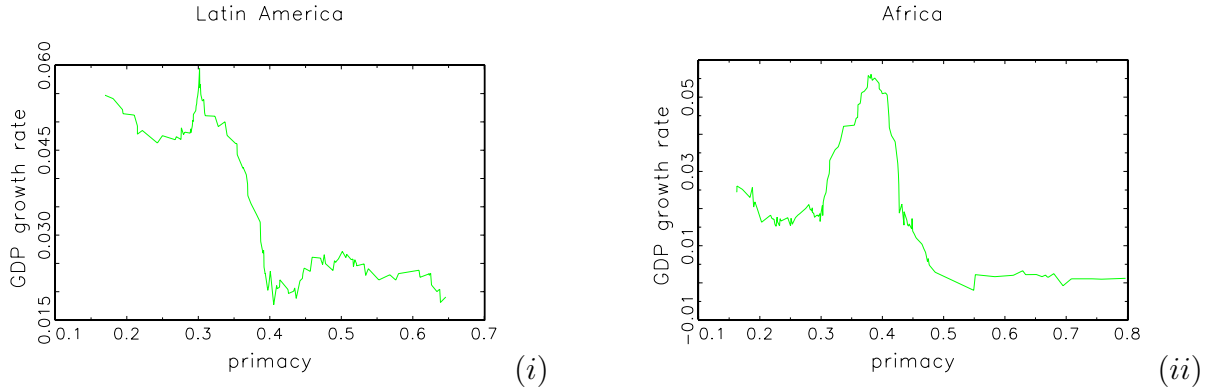


Figure 2: Latin America and African non-parametric curves. OLS procedure

0.55. After that value it exhibits a globally flat pattern. This picture suggests that African countries that have levels of primacy below that maximum have not exhausted agglomeration economies yielded by their primate cities. Conversely, African countries with levels of primacy greater than that maximum may be expected to have congestion, pollution and other centrifugal forces overwhelming agglomeration economies.

Results from tests of equality of non-parametric regressions functions estimated by OLS procedures signal heterogeneity in the overall sample, confirming the impression yielded by the previous pictures (Table 5). Deeper inspection indicates however that this heterogeneity lies mostly in the opposition between developing countries and developed ones, and between Europe and other regions of the World. For regions pertaining to the developing World there is no evidence of significantly different patterns. So there is some gap between the result of those statistical tests and the impression provided by the pictures.

Table 5: Tests of equality of non-parametric regression function

	P-value
Overall sample	0.004**
Developing VS Developed countries	0.000**
Europe VS Africa	0.000**
Asia VS Africa	0.323
Latin America VS Africa	0.433

Obviously IV procedures are expected to be the most reliable, yet they provide results that are very similar to the previous curves.¹⁰ The non-parametric curve obtained from European countries is globally increasing (Figure 3) suggesting as before that urban development in that region is well balanced and efficient.

While ‘overconcentration’¹¹ does not appear to be a pertinent description of Europe, the picture from the developing countries appears to be more mitigated. Inspection of non-parametric curves for Asia and Latin America uncovers as before the existence of urban concentration traps. Such trap is much more significant for Latin America where a significant range of urban primacy appears to be associated with negative economic growth.¹² This is consistent with the peculiar spatial distribution prevailing in Latin America. This continent is well known for the loose integration of several of its very vast regions. This has induced a spatial configuration with a strong urban concentration and sharp contrasts between on one hand fast growing metropolitan areas and vast abandoned rural regions. This result is also consistent with the failure of Latin America development strategy based on import substitution and comforts the idea that Latin America is the region of the world where primacy issues are the most involved.

Asia faces a similar urban concentration trap. Yet, the negative impact of primacy on its economic development seems less dramatic, most of the range of primacy variation being associated with a positive economic growth. The picture from Black Africa confirms previous results as well. Black Africa remains the only geographical area where the pat-

¹⁰Yet, because of the loss of degrees of freedom, the Europe non-parametric curves obtained for the instrumental variables procedure seem to be very imprecise. Thus, in Appendix F devoted to confidence intervals, we replace them by non-parametric curves estimated for developed countries.

¹¹The term is understood here as the prevalence of excessive urban concentrations

¹²Argentina, Chile, Dominican Republic, Guatemala, Nicaragua, Peru and El Salvador are the countries of that region where primacy seems to have the most detrimental impact on economic growth.

tern of the relationship between growth and primacy is similar to the prediction of the Williamson hypothesis. Inspection of the African non-parametric curve indicates a maximum at about 37%. Thus, urban concentration of countries having a primacy below that level is not detrimental to economic growth. The lopsided spatial distribution of several African countries does not seem to be an obstacle to their economic development as they do not appear to have exhausted their agglomeration economies. Mali, Chad and Zimbabwe are the countries that are the closest to that optimum. Conversely, SSA countries with a primacy higher than 0.45 - like Angola, Congo, Guinea, Mozambique, Senegal - have the worst economic growth rates.

While differences appear from the comparison of the patterns of different regions, tests of equality of non-parametric functions don't signal any difference between the different non-parametric curve at the 5% signification level. This contrasts singularly with the result obtained with the OLS estimation procedures. This striking contrast may be caused by the loss of degrees of freedom implied by the use of lagged variables as instruments in IV estimation methods.

Table 6: Tests of equality of non-parametric regression function

	P-value
Overall sample	0.149
Developing VS Developed countries	0.346
Europe VS Africa	0.193
Asia VS Africa	0.123
Latin America VS Africa	0.365

3.2.3 Productivity model with urban density

Parametric estimates

Table 7 shows results of a productivity model including urban density as the urban concentration measure. As for primacy estimations, OLS estimation procedure yields a capital coefficient that is much higher than what is commonly assumed in the literature. Conversely to primacy estimations, OLS estimate of the human capital coefficient is significant.

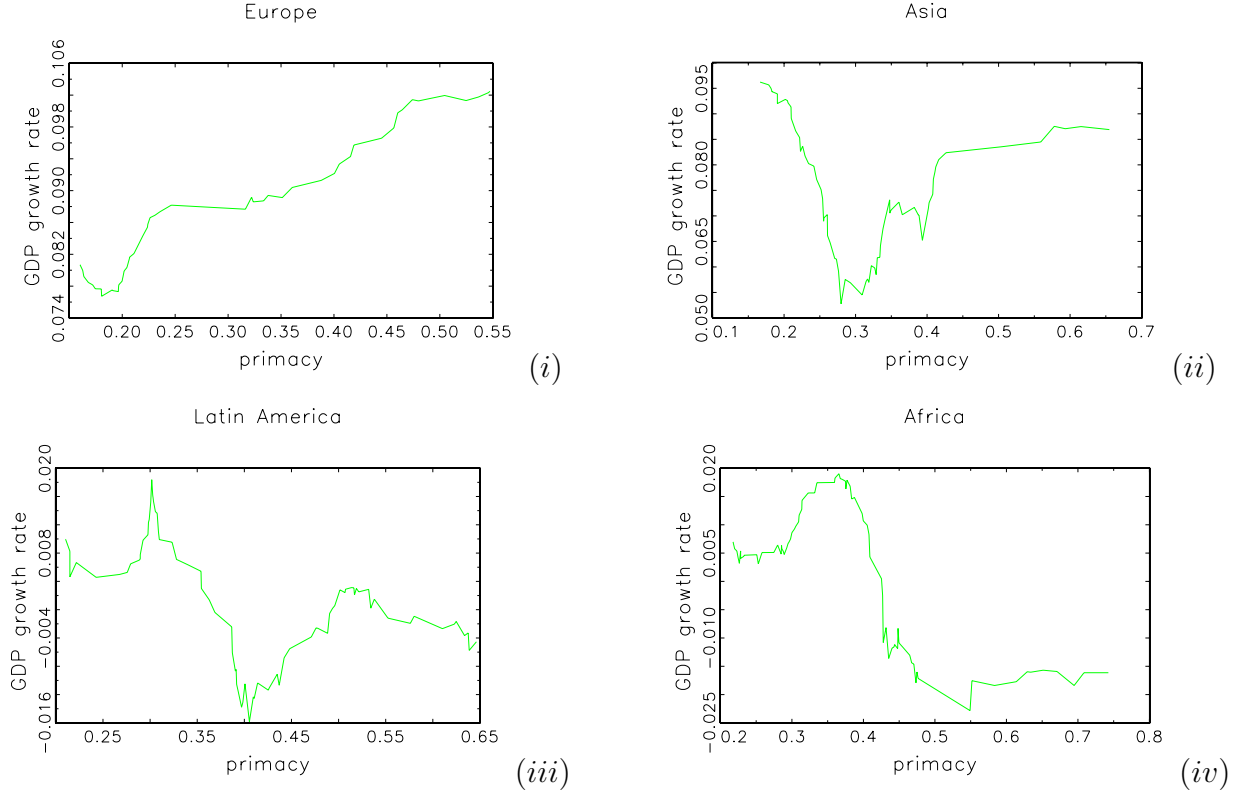


Figure 3: Non-parametric curves. IV procedure

Table 7: Estimation of a productivity growth equation with urban density. Dependent variable is $\Delta \ln(Y_{(t)}/L_{(t)})$.

	(1) OLS	(2) IV
$\Delta \ln(K_{(t)}/L_{(t)})$	0.5148**	0.2802**
Human capital	0.0192**	0.0207**
Years effects	Yes	Yes
N[countries]	582[97]	388[97]

However, while its value is a little higher than the ones obtained by primacy estimation procedures, it is still lower than Henderson' (2003) GMM estimation.

Instrumental variables estimation with the urban density variable corrects the capital coefficient which reaches a value that is even lower than what is generally assumed by the literature. Yet, IV estimate of the human capital coefficient is only slightly higher than the OLS one, indicating as before that the human capital coefficient is quite unaffected by any endogeneity bias.

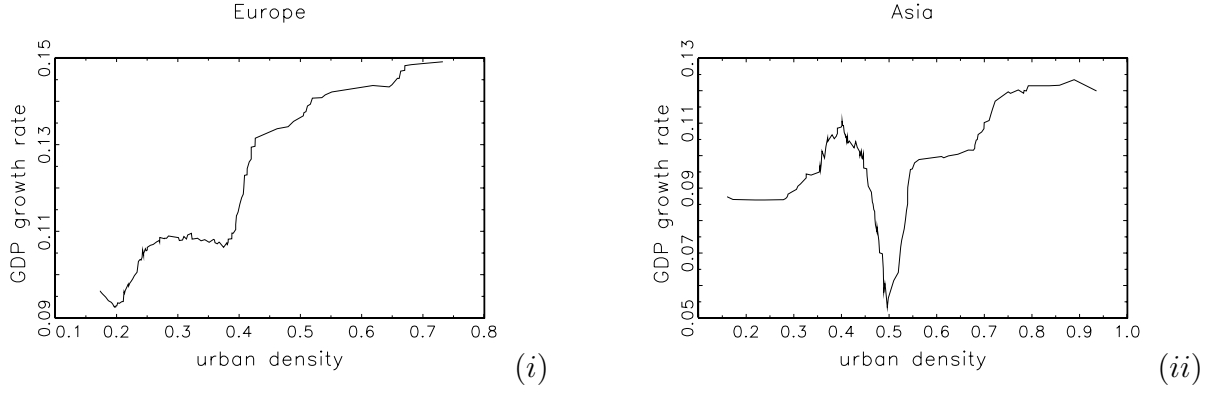


Figure 4: European and Asian non-parametric curves. OLS procedure

Non-parametric curves

From OLS estimation we find that the pattern exhibited by the non-parametric curve for Europe (Figure 4) is as previously globally increasing. The shape is very similar to the one exhibited by the primacy non-parametric curve obtained under OLS estimation. Indeed, except for values of urban density below 0.2 and for values lying in the 0.27-0.39 range where the curve is flat, the curve exhibits a positive slope which is the highest in the 0.20-0.27 and the 0.39-0.42 ranges. This confirms the previous result that in Europe urban concentration seems to be associated to economic growth for most of its variation range.

The previous findings appear to be robust for Asia and Latin America as well. A urban density trap is clearly visible in the non-parametric curve drawn for Asia and Latin America in OLS estimation procedure. As previously, Latin America non-parametric curve is the least smooth. Results for Africa diverge from previous findings. While a inverted U-shape was clearly emerging for Africa in primacy non-parametric curves, this is clearly not the case for African OLS urban density non-parametric curve since it displays a succession of local maxima with a globally decreasing shape. This very irregular pattern does not ease any global interpretation of the impact of urban concentration of the economic growth of countries of this region of the world, urban density may raise or lower economic growth depending on the specific range of variation that we consider.

Tests of equality of non-parametric curves comfort at first sight this picture of diverging patterns across regions. While for the overall sample, heterogeneity in the non-parametric curves does not appear significant, tests of equality signal significant difference between developed and developing countries. However, conversely to differences suggested by the graphs of non-parametric curves, those tests fail to indicate any significant difference within the developing world between Africa and the Latin American and Asian continents.

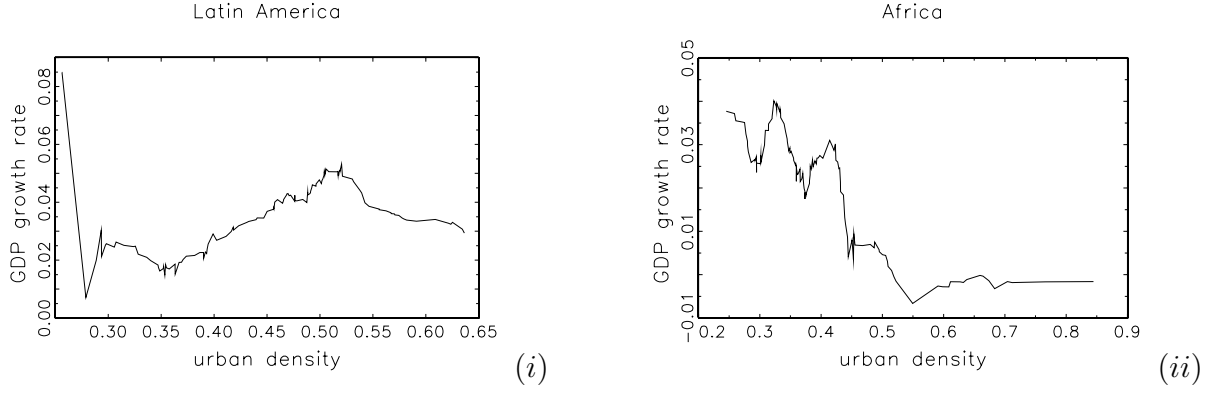


Figure 5: Latin America and African non-parametric curves. OLS procedure

Table 8: Tests of equality of non-parametric regression function

	P-value
Overall sample	0.099*
Developing VS Developed countries	0.054*
Europe VS Africa	0.007**
Asia VS Africa	0.114
Latin America VS Africa	0.142

Except for Africa IV estimation provides patterns of non-parametric curve that are very close to those obtained by OLS procedures. IV procedures even yield an Europe non-parametric curve that has an increasing shape for all the variation range of the urban density variable (Figure 6). Africa non-parametric curve exhibits a globally decreasing shape (Figure 7). This result suggests that countries of that region have not the economic infrastructure that may support a high share of population in big cities. Furthermore, as average urban density for the Black Africa amounts to about 0.39, such a non-parametric curve clearly indicates that Sub-Saharan Africa is overurbanized. But as this result diverges from previous findings with primacy non-parametric curves, this outlines more seriously that in some regions those two indicators of urban concentration are poorly correlated.

As before tests of equality of non-parametric regression curves yield contrasted results. While tests applied to the overall sample confirm the impression yields by the graphs of diverging patterns, they fail to confirm that African and Latin America non-parametric curves are significantly different.

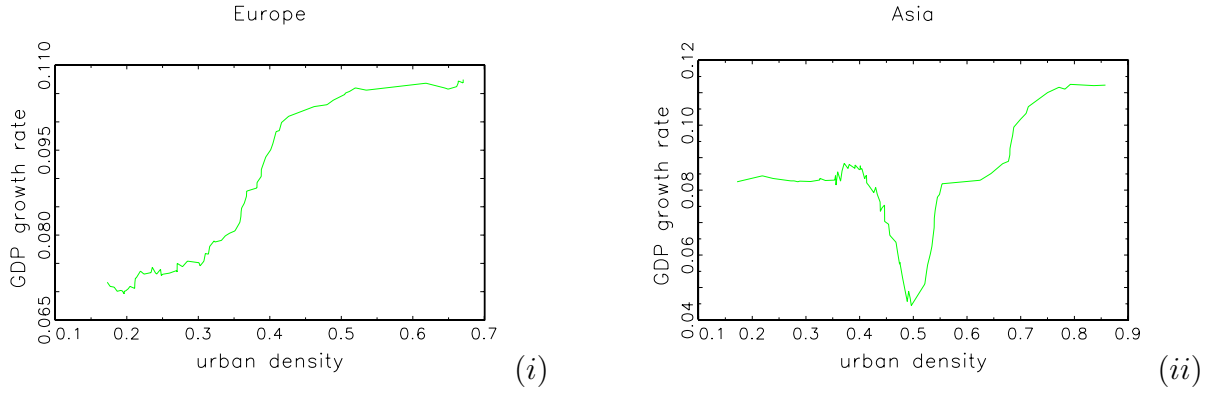


Figure 6: European and Asian non-parametric curves. IV procedure

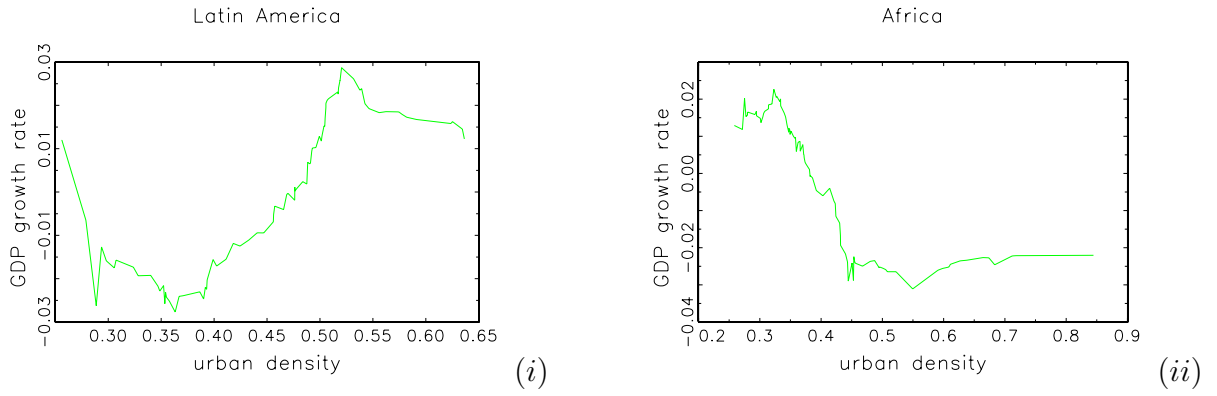


Figure 7: Latin America and African non-parametric curves. IV procedure

Table 9: Tests of equality of non-parametric regression function

	P-value
Overall sample	0.011**
Developing VS Developed countries	0.136
Europe VS Africa	0.021**
Asia VS Africa	0.031**
Latin America VS Africa	0.431

4 Conclusion

In a pretty old statement Hoselitz (1955) raised the claim that there is a contrast between urban development in developed nations and in developing countries: while in the former group there is an intimate connection between the economic demands for labor exerted by progressive accumulation of capital in urban industry and the growth of urban centers, this is not the case in many of the underdeveloped countries in Asia, Africa, and Latin America, where a number of other reasons seem to have induced rural-urban migration. That statement about the possible qualitative difference of the urban development in different groups of countries is at the heart of the issue addressed by this paper.

Pointing out the heterogeneity between regions, our results back Hoselitz statement: urban concentration has a positive impact on economic growth in Europe, dummy traps prevails for Latin America and Asia, while Africa non-parametric curve differs depending of the measure of urban concentration considered. It exhibits a bell-shaped pattern with primacy but a curve with a decreasing slope when urban density is considered. In most cases tests of equality of non-parametric regression functions confirm the heterogeneity hypothesis.

Therefore, no general relationship between urban concentration and economic growth appears as credible. Any attempt to assess the impact of the spatial distribution on economic development should be addressed to groups of homogeneous countries. In this respect even the grouping we use for this empirical research may be improved. The poor smoothness of some of the non-parametric curves estimated, especially for Asia and Latin America may indicate that those groups are hardly homogeneous.

Moreover, Africa non-parametric curve does not seem to display a monotonically decreasing pattern with respect to urban primacy. Therefore, we may not discard the importance of agglomerations economies in Africa urbanization process. Those agglomerations economies will play an important role in subsequent papers devoted to the modeling of the impact of respectively political factors and locational (dis)advantages on urban concentration.

References

- [1] Ades, A. and E.L. Glaeser (1995) Trade and circuses: Explaining urban giants, *Quarterly Journal of Economics*, 110, 195-227.
- [2] Alonso, W. (1980) Five Bell Shapes in Development, *Papers of the Regional Science Association*, 45, 5-16.
- [3] Bairoch, P. (1985) *De Jericho à Mexico: Villes et économie dans l'histoire*. Paris: Gallimard. English translation *Cities and Economic Development. From the Dawn of History to the Present* (1988), London: Mansell Publishing Ltd.
- [4] Barro, R. and J.-W. Lee (2001) International data on educational attainment: updates and implications, *Oxford Economic Papers* 3, 541-563.
- [5] Bauwens, L. and H.K. Van Dijk (1990) Bayesian limited information analysis revisited. In: Gabszewicz, J.J., J.F Richard and L.A. Wolsey (Eds.) *Economic Decision Making: Games, Econometrics and Optimisation*, North-Holland, pp. 385-424.
- [6] Bertinelli, L. and E. Strobl (2003) Urbanization, Urban Concentration and Economic Growth in Developing Countries, CORE Discussion Paper #2003/76.
- [7] Black, D. and J.V. Henderson (1999) A Theory of Urban Growth, *Journal of Political Economy* 107, 252-284.
- [8] Dareshwar, A. and V. Nehru (1993) A New Database on Physical Capital Stock: Sources, Methodology and Results, *Revista de Analisis Economico* 8, 37-59.
- [9] Davis, J.C. and J.V. Henderson (2003) Evidence on the political economy of the urbanization process, *Journal of Urban Economics* 53, 98-125.
- [10] Dreze, J.H. and J.F. Richard (1983) Bayesian Analysis of Simultaneous Equation Systems. In: Griliches Z., M.D. Intriligator (eds.) *Handbook of Econometrics*, North-Holland: Elsevier Science Publishers, pp. 517 - 598.
- [11] Duranton, G. and D. Puga (2001) Nursery Cities, *American Economic Review* 91, 1457-1477.
- [12] Durlauf, S. and D. Quah (1998) The New Empirics of Economic Growth, NBER Working Papers #6422.

- [13] Eaton, J. and Z. Eckstein (1997) Cities and Growth: Theory and Evidence from France and Japan, *Regional Science & Urban Economics* 27, 443-474.
- [14] El-Shakhs, S. (1972) Development, Primacy, and Systems of Cities, *Journal of Developing Areas* 7, 11-36.
- [15] Fay, M. and C. Opal (2000) Urbanization without growth: A not so uncommon phenomenon, World Bank Working Paper #2412.
- [16] Fujita, M., P.R. Krugman and A.J. Venables (1999) *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.
- [17] Grossman, G.M., and E. Helpman (1991) *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.
- [18] Hall, R.E. and C.I. Jones (1999) Why Do Some Countries Produce So Much More Output Per Worker than Others, *Quarterly Journal of Economics* 83-116.
- [19] Henderson, J.V. (1974) The Sizes and Types of Cities, *American Economic Review* 64, 640-656.
- [20] Henderson, J.V., T. Lee, and Y.J. Lee (2001) Scale Externalities in Korea, *Journal of Urban Economics* 49, 479-504.
- [21] Henderson, J. V. (2003) The Urbanization Process and Economic Growth: The So-What Question, *Journal of Economic Growth* 8, 47-71.
- [22] Hoselitz, B. F. (1955) Generative and Parasitic Cities, *Economic Development and Cultural Change* 3, 278-294.
- [23] Junius, K. (1999) Primacy and Economic Development: Bell shaped or parallel growth of cities? *Journal of Economic Development* 24, 1-22.
- [24] King, G., J. Honaker, A. Joseph, and K. Scheve (2001) Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation, *American Political Science Review* 95, 49-69.
- [25] Koop, G. and D.J. Poirier (2004) Bayesian variants of some classical semiparametric regression techniques, *Journal of Econometrics* 123, 259-282.

- [26] Magrini, S. (2004) Regional (di)convergence. In: Henderson, J.V. and J.F. Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4, North-Holland: Elsevier Science Publishers, pp. 2741-2796.
- [27] Mazumdar, J. G. (1987) Rural-urban migration in developing countries. In: Mills, E.S. (Ed.) *Handbook of Regional and Urban Economics*, vol. 2, North-Holland: Elsevier Science Publishers, pp. 1097-1128.
- [28] Mera, K. (1973) On the Urban Agglomeration and Economic Efficiency?, *Economic Development and Cultural Change* 21, 309-324.
- [29] Moomaw, R.L. and A.M. Shatter (1996) Urbanization and economic development: a bias toward large cities?, *Journal of Urban Economics* 40, 13-37.
- [30] Mutlu, S. (1989) Urban Concentration and Primacy Revisited: An Analysis and Some Policy Conclusions, *Economic Development and Cultural Change* 37, 611-639.
- [31] Richardson, H.W. and G. Schwartz (1988) Economic Development, Population, and Primacy, *Regional Studies* 22, 467-475.
- [32] Schafer, J.L. (1997) *Analysis of incomplete multivariate data*. Boca Raton: Chapman & Hall/CRC.
- [33] Schimert, J., J.L. Schafer, T. Hesterberg, C. Fraley and D.B. Clarkson (2001) *Analysing data with missing values in S-Plus*. Seattle: Insightful.
- [34] Temple, J. (1999) The New Growth Evidence, *Journal of Economic Literature* 37, 112-156.
- [35] Todaro, M.P. (1969) A Model of Labor Migration and Urban Unemployment in Less Developed Countries, *American Economic Review* 59, 138-148.
- [36] Wheaton, W. and H. Shishido (1981) Urban Concentration, Agglomeration Economies, and the Level of Economic Development *Economic Development and Cultural Change* 30, 17-30.
- [37] Williamson, J. G. (1965) Regional Inequality and the Process of National Development: A Description of the Patterns *Economic Development and Cultural Change* 13, 3-84.

- [38] Williamson, J. G. (1987) Migration and Urbanization. In: Chenery, H. and T.N. Srinivasan (eds.) *Handbook of Development Economics*, vol. 1, North-Holland: Elsevier Science Publishers, pp. 424-465.
- [39] Yatchew, A. (1998) Nonparametric Regression Techniques in Economics, *Journal of Economic Literature* 36, 669-721.
- [40] Yatchew, A. (1999) An elementary nonparametric differencing test of equality of regression functions, *Economic Letters* 62, 271-278.
- [41] Yatchew, A. (2003) *Semiparametric regression for the applied econometrician*. Cambridge: Cambridge university press.
- [42] Yuki, K. (2007) Urbanization, informal sector, and development *Journal of Development Economics* 84, 76-103.
- [43] Zellner, A., L. Bauwens and H.K. Van Dijk, (1988) Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods *Journal of Econometrics* 38, 39-72.

Appendix A: Sample and Data Sources

The dataset consists of 97 countries.

Country list: Afghanistan, Algeria, Angola, Argentina, Australia, Austria, Belgium, Burkina Faso, Bangladesh, Bolivia, Brazil, Bulgaria, Canada, Chad, Chile, China, Côte d'Ivoire, Cameroon, Congo (Democratic Republic), Congo (Republic), Colombia, Costa Rica, Cuba, Czechoslovakia, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Ethiopia, Finland, France, Ghana, Greece, Guatemala, Guinea, Haiti, Honduras, Hungary, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Japan, Jordan, Kenya, Libya, Madagascar, Malaysia, Mali, Mexico, Morocco, Mozambique, Myanmar, Netherlands, New Zealand, Nicaragua, Nigeria, North Korea, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Romania, Saudi Arabia, Senegal, Somalia, South Africa, South Korea, Spain, Sudan, Sweden, Switzerland, Syria, Tanzania, Thailand, Tunisia, Turkey, Uganda, Union of Soviet Socialist Republics, United Arab Emirates, United Kingdom, USA, Uruguay, Venezuela, Vietnam, West Germany, Yugoslavia, Zambia, Zimbabwe.

Table 10: Data Sources

Data	Sources
GDP/Capita Capital/Worker	Dhareshwar, A. and V. Nehru (1993), A New Database on Physical Capital Stock: Sources, Methodology and Results, Revista de Analisis Economico, 8, pp.37-59 http://www.worldbank.org/research/growth/ddnehdha.htm
Urban density Primacy	National urban population. UN World Urbanization Prospects CD-rom, File 10 POP/DB/WUP/Rev.2003/2/F10
Human capital	Measured by average schooling years in the male population aged 25 and over. Stems from Henderson (2003). Figures were obtained from Barro, R. and J.-W. Lee (2001) and from Census and survey figures primarily retrieved from UNESCO Statistical Yearbooks and UN Demographic Yearbooks. Remaining values are estimated using UNESCO school enrollment data and a perpetual inventory method. The data are not adjusted for quality of education day or or length of school year.

Appendix B: Descriptive statistics

Table 11: Descriptive statistics

Variables	Min	Max	Mean	Std Dev
Human capital	0.020	6.220	1.158	1.050
Primacy	0.044	0.797	0.312	0.149
Urban density	0.099	0.935	0.420	0.148
ln(Capital per worker)	6.994	11.706	9.622	1.176
ln(GDP per capita)	3.201	12.114	7.669	1.991

Table 12: Proportion of missing data by region

Region	Missing values	Sample size	%Missing
Asia	53	161	32.92
Europe	42	147	28.57
Latin America	14	147	9.52
North Africa	17	35	48.57
North America	0	14	0.00
Oceania	0	14	0.00
Sub Saharan Africa	78	161	48.57

Appendix C: Imputation of covariates

For an arbitrary pattern of missing data parameter estimates cannot be obtained in closed form. Therefore, we resort to a data augmentation algorithm which implies iterative computations. To ease those computations, it is useful to group the rows of the covariates by their missingness pattern.

Following Schafer (1997), we may index the missingness patterns by $s = 1, 2, \dots, S$, where S is the number of unique patterns prevailing in the covariates data matrix.¹³ For a given data matrix X of dimension $n \times p$, with $n = NT$ let's define R as an $S \times p$ matrix of binary indicators with typical elements r_{sj} , where

$$r_{sj} = \begin{cases} 1 & \text{if } X_j \text{ is observed in pattern } s, \\ 0 & \text{if } X_j \text{ is missing in pattern } s. \end{cases} \quad (33)$$

Table 13: Matrix of missingness patterns associated with X .

	X_1	X_2	X_3	\dots	X_p
patterns s=1	1	1	1		1
2	0	1	1		1
.	1	0	1		1
.	0	0	1		1
.	1	1	0		1
.
.
.
.	0	1	0		0
S	1	0	0		0

Table 13 shows the typical matrix R . For each missingness pattern s , let $\mathcal{O}(s)$ and $\mathcal{M}(s)$ denote the subsets of the columns labels $\{1, 2, \dots, p\}$ corresponding to variables that are observed and missing, then we have respectively,

$$\begin{aligned} \mathcal{O}(s) &= \{j : r_{sj} = 1\} \\ \mathcal{M}(s) &= \{j : r_{sj} = 0\} \end{aligned} \quad (34)$$

Finally we denote by $\mathcal{I}(s)$ the subset of $\{1, 2, \dots, n\}$ corresponding to rows of the data matrix exhibiting missingness pattern s .

¹³In this section our presentation will closely follow Schafer (1997)

The I-step

Since we assume that the rows of a data matrix are conditionally independent given θ simulation of (23) is carried out by drawing

$$x_{i(mis)}^{(t+1)} \sim P(x_{i(mis)} | x_{i(obs)}, \theta^{(t)}), \quad (35)$$

independently for $i = 1, 2, \dots, n$. For a given row i in missingness pattern s the conditional distribution of $x_{i(mis)}$ given $x_{i(obs)}$ and θ is multivariate normal with means

$$E(x_{ij} | X_{obs}, \theta) = a_{0j} + \sum_{k \in \mathcal{O}(s)} a_{kj} x_{ik} \quad (36)$$

and covariances

$$Cov(x_{ij}, x_{ik} | X_{obs}, \theta) = a_{jk} \quad (37)$$

with $j, k \in \mathcal{M}(s)$, and a_{jk} denoting an element of the matrix

$$A = SWP[\mathcal{O}(s)] \theta \quad (38)$$

$SWP[\]$ denotes the sweep operator. When applied to the parameters of a multivariate normal model it converts a variable from a response to a predictor. Considering $z \sim \mathcal{N}(\mu, \Sigma)$ a random vector of variables partitioned as $z^T = (z_1^T, z_2^T)$ with p_1 the length of z_1 , $SWP[1, \dots, p_1] \theta$ converts the parameters matrix θ

$$\theta = \begin{bmatrix} -1 & \mu^T \\ \mu & \Sigma \end{bmatrix} = \begin{bmatrix} -1 & \mu_1^T & \mu_2^T \\ \mu_1 & \Sigma_{11} & \Sigma_{12} \\ \mu_2 & \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad (39)$$

containing parameters of the marginal distributions of z_1 and z_2 to a matrix containing parameters of the conditional distribution of z_2 given z_1

$$SWP[1, \dots, p_1] \theta = \begin{bmatrix} -1 - \mu_1^T \Sigma_{11}^{-1} \mu_1 & \mu_1^T \Sigma_{11}^{-1} & \mu_2^T - \mu_1^T \Sigma_{11}^{-1} \Sigma_{12} \\ \Sigma_{11}^{-1} \mu_1 & -\Sigma_{11}^{-1} & \Sigma_{11}^{-1} \Sigma_{12} \\ \mu_2 - \Sigma_{12} \Sigma_{11}^{-1} \mu_1 & \Sigma_{21} \Sigma_{11}^{-1} & \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \end{bmatrix} \quad (40)$$

$$= \begin{bmatrix} -1 - \mu_1^T \Sigma_{11}^{-1} \mu_1 & \mu_1^T \Sigma_{11}^{-1} & \alpha_{2.1}^T \\ \Sigma_{11}^{-1} \mu_1 & -\Sigma_{11}^{-1} & B_{2.1}^T \\ \alpha_{2.1} & B_{2.1} & \Sigma_{22.1} \end{bmatrix} \quad (41)$$

with

$$\alpha_{2.1} = \mu_2 - \Sigma_{12} \Sigma_{11}^{-1} \mu_1 \quad (42)$$

$$B_{2.1} = \Sigma_{21} \Sigma_{11}^{-1} \quad (43)$$

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \quad (44)$$

The P-step

Assuming that no prior information is available concerning θ , the complete data posterior is a normal-inverted Wishart distribution. Therefore, the P-step will consist merely of the following simulation

$$\mu|\Sigma, X \sim N(\bar{x}, n^{-1}\Sigma) \quad (45)$$

$$\Sigma|X \sim W^{-1}(n-1, (nS)^{-1}) \quad (46)$$

where \bar{x} and S denote respectively sample mean and sample covariance matrix.

Appendix D: Convergence diagnosis

Each iteration of our overall estimation algorithm will consist in a completion of the imputation algorithm which will imply numerous iterations to achieve convergence and one iteration of each step of the algorithm for Bayesian estimation and imputation of missing values of the dependent variable. Therefore, if respectively N_I and N_E iterations are needed for the imputation and the Bayesian estimation algorithms the overall algorithm will perform $N_I \times N_E$ imputations and N_E parameter estimations.

To obtain reliable imputations or parameter estimates we have to run those Gibbs Sampler algorithms enough times to allow the algorithm to converge to the posterior distributions. Several procedures and statistics are available to assess convergence:

- Times series plot and autocorrelations: plotting iterates of components of θ is a quick and easy way to assess convergence. In case of fast convergence plots of iterates show no discernible trends; they resemble horizontal bands indicating a low ratio of noise to signal. For imputation algorithm it corresponds to situations where the fraction of missing data is moderate. Conversely, when the fraction of missing information is high, long-term trends and high serial correlation are likely to show up and the algorithm converges slowly. Another way to assess convergence is to investigate the relationship between iterates at time (t) and at time $(t+1)$. This may be done through the analysis of the autocorrelation function. If autocorrelations between draws die out very quickly, the convergence is fast. Conversely, if correlations are still high beyond 10 iterations draws display a high degree of serial dependence and convergence is slow.
- Geweke's test statistic: compares the estimate \bar{g}_A of a posterior mean from the first draws with the estimate from the last draws \bar{g}_B . If the two subsamples (of size n_A and n_B) are well separated (i.e. there are many observations between them), they should be independent. The statistic, normally distributed if n is large and the chain has converged, is

$$Z = \frac{\bar{g}_A - \bar{g}_B}{nse_A^2 + nse_B^2}$$

where nse_A and nse_B represent numerical standard errors of each subsample

- Standardized CUMSUM statistic: the standardized CUMSUM for θ is:

$$CS_t = \left(\frac{1}{t} \sum_{i=1}^t \theta^i - m_\theta \right) / s_\theta$$

where m_θ and s_θ are the MC sample mean and standard deviation of the n draws. If the MCMC sampler converges, the graph of CS_t against t should converge smoothly to zero. On the contrary, long and regular excursions away from zero are an indication of the absence of convergence. A value of 0.05 for a CUMSUM after t draws means that the estimate of the posterior expectation diverges from the final estimate (after n draws) by 5 per cent in units of the final estimate of the posterior standard deviation.

Here are the graphics obtained for the autocorrelations, the standardized CUMSUM statistic (convergence graphs) and the sequence of draws for respectively 2000 and 5000 iterations:

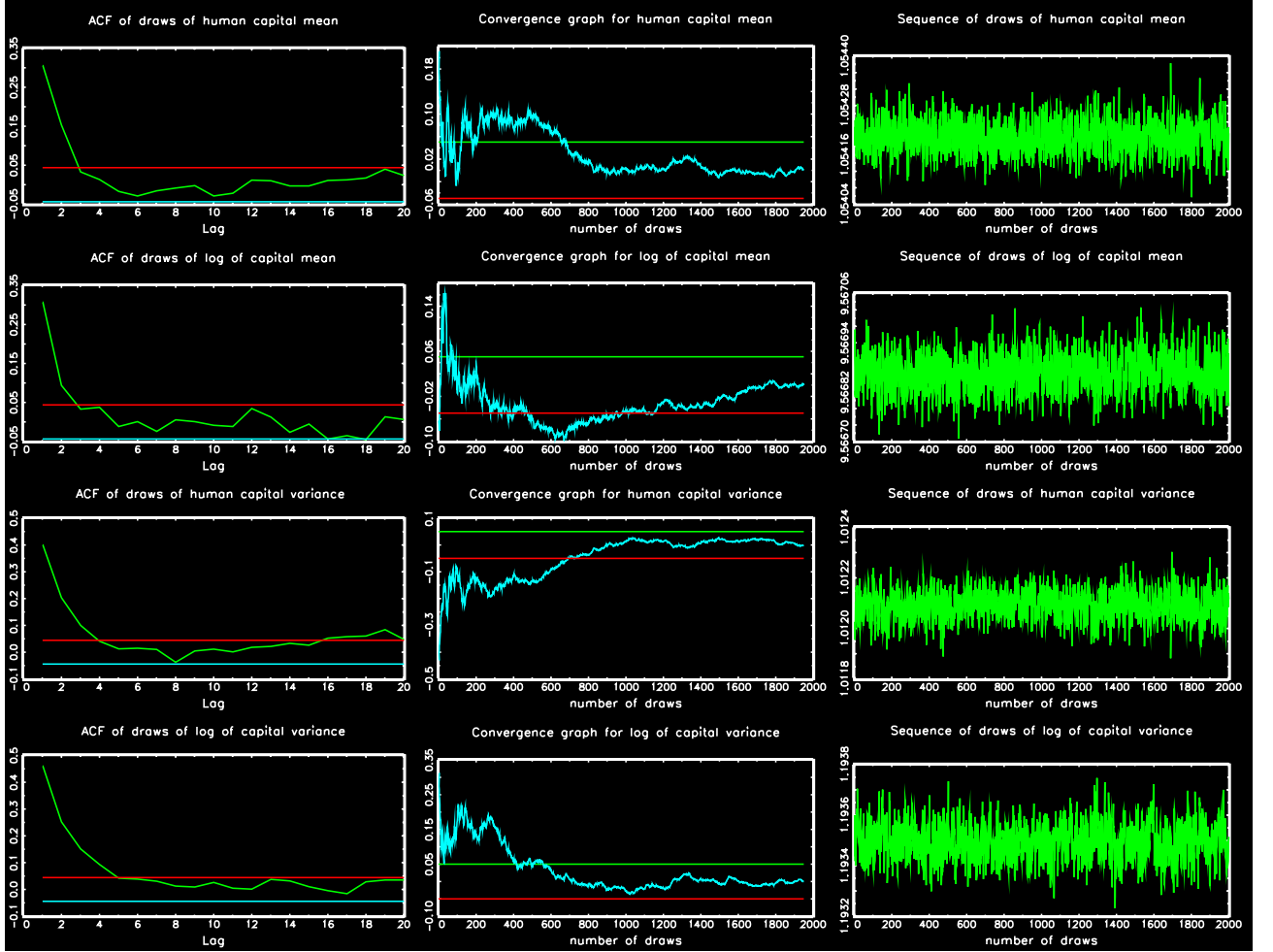


Figure 8: Convergence Diagnosis for 2000 iterations

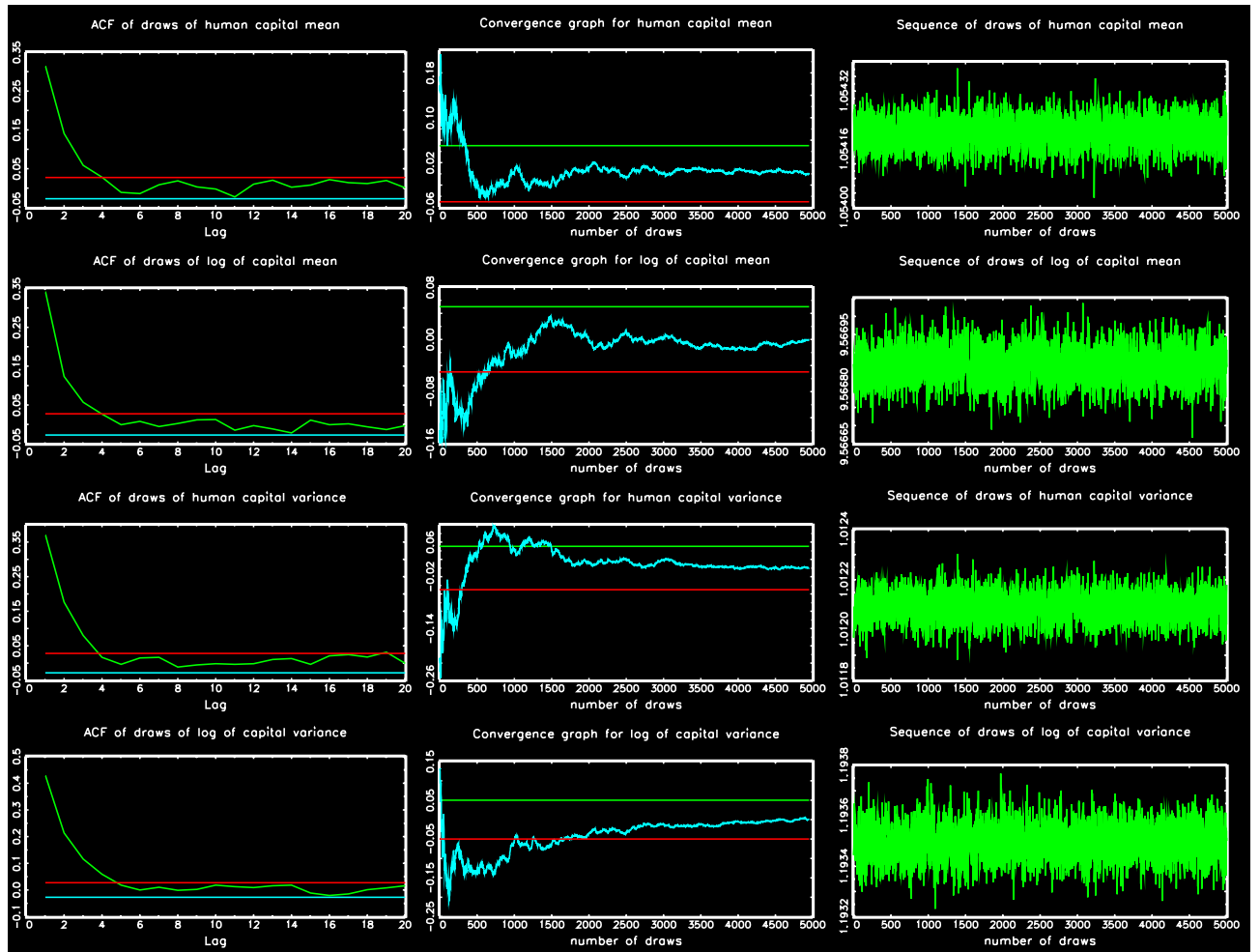


Figure 9: Convergence Diagnosis for 5000 iterations

Appendix E: Algorithm Pseudocode

In order to summarize our estimation procedure, we give hereafter our algorithm pseudocode, where D represents the data matrix. All the estimation and imputations procedures were implemented in Gauss.

```

for  $i := 1$  to  $N_E$  do
  for  $j := 1$  to  $N_I$ 
    draw of  $X_{mis}^{t+1} \sim P(X_{mis}/X_{obs}, \theta^t)$ 
    draw of  $\theta^{t+1} \sim P(\theta/X_{obs}, X_{mis}^{t+1})$ 
  endfor

  if not endogeneity
    draw of  $y_{mis}^{t+1}|y_{obs}, X, \beta^t, \sigma_t^2 \sim \mathcal{N}(X\beta^t, \sigma_t^2)$ 
    draw of  $\beta^{t+1}, \sigma_{t+1}^2|y_{obs}, X, y_{mis} \sim NIG(\hat{\beta}, X'X, s, n - k - 2)$ 

  else
    draw of  $y_{mis}^{t+1}|y_{obs}, X, \delta_2^{t+1}, \sigma_t^2 \sim \mathcal{N}(X\delta_2^{t+1}, \sigma_t^2)$ 
    draw of  $\delta_2^{t+1} \sim p(\delta_2|\Pi^t, D)$ 
    draw of  $\Pi^{t+1} \sim p(\Pi|\delta_2^{t+1}, D)$ 
  endfor
endfor

```

Appendix F: Confidence intervals

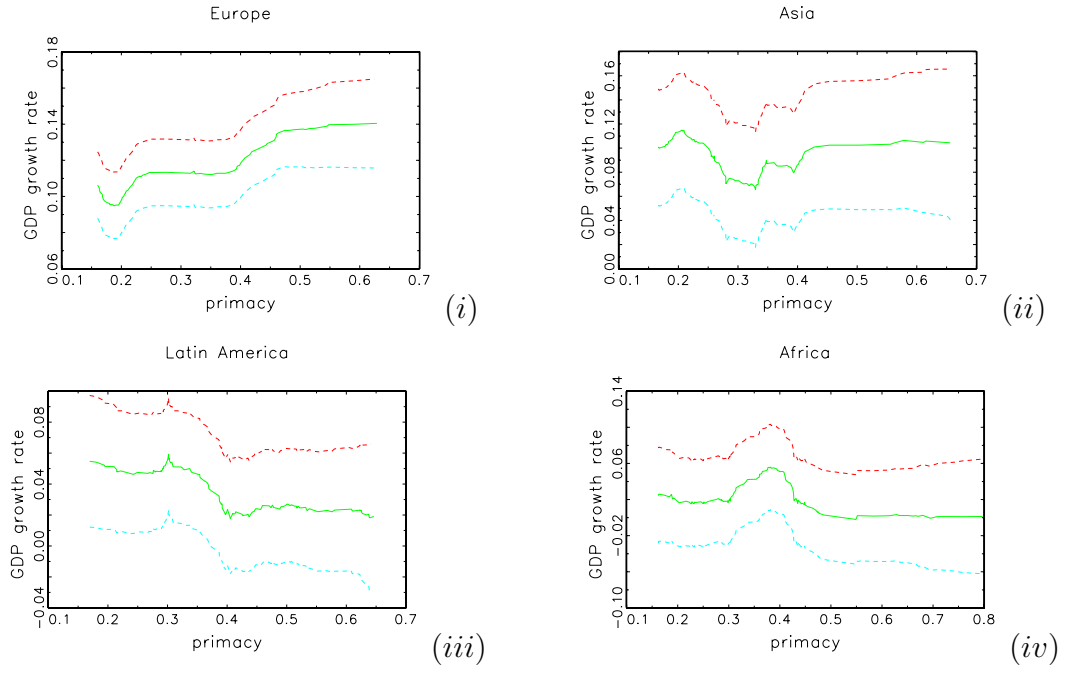


Figure 10: Non-parametric curves with primacy. OLS procedure

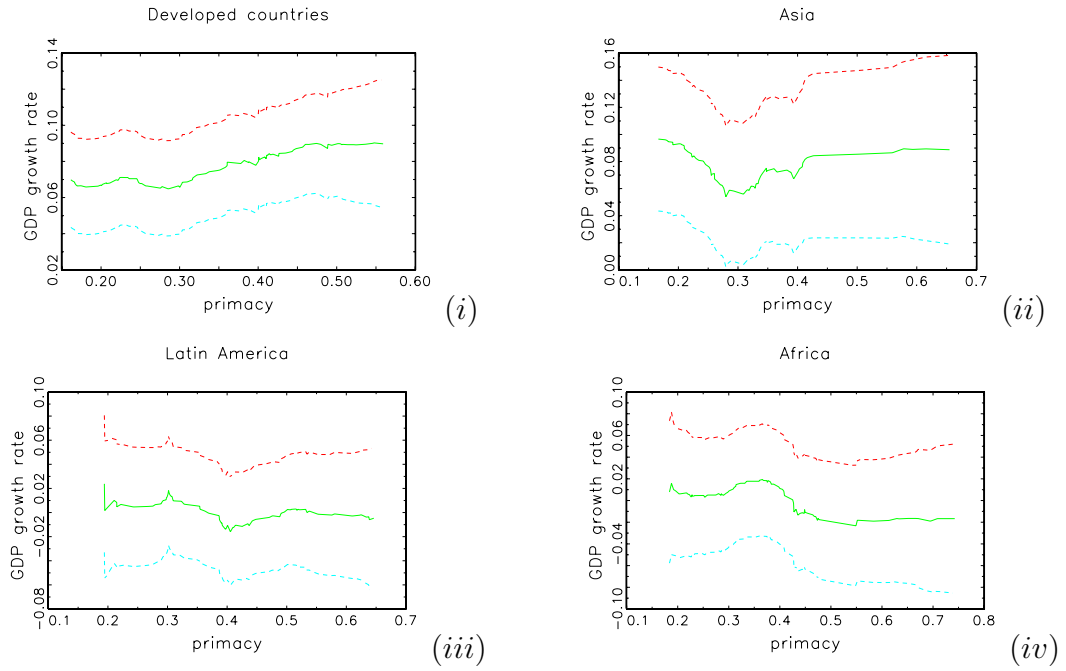


Figure 11: Non-parametric curves with primacy. IV procedure

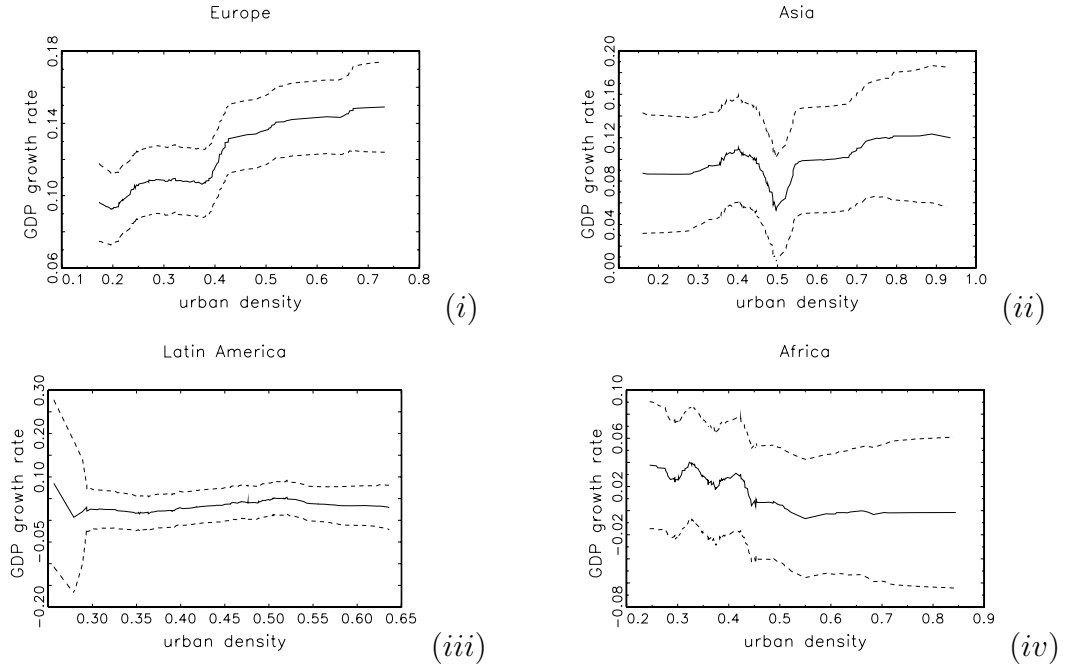


Figure 12: Non-parametric curves with urban density. OLS procedure

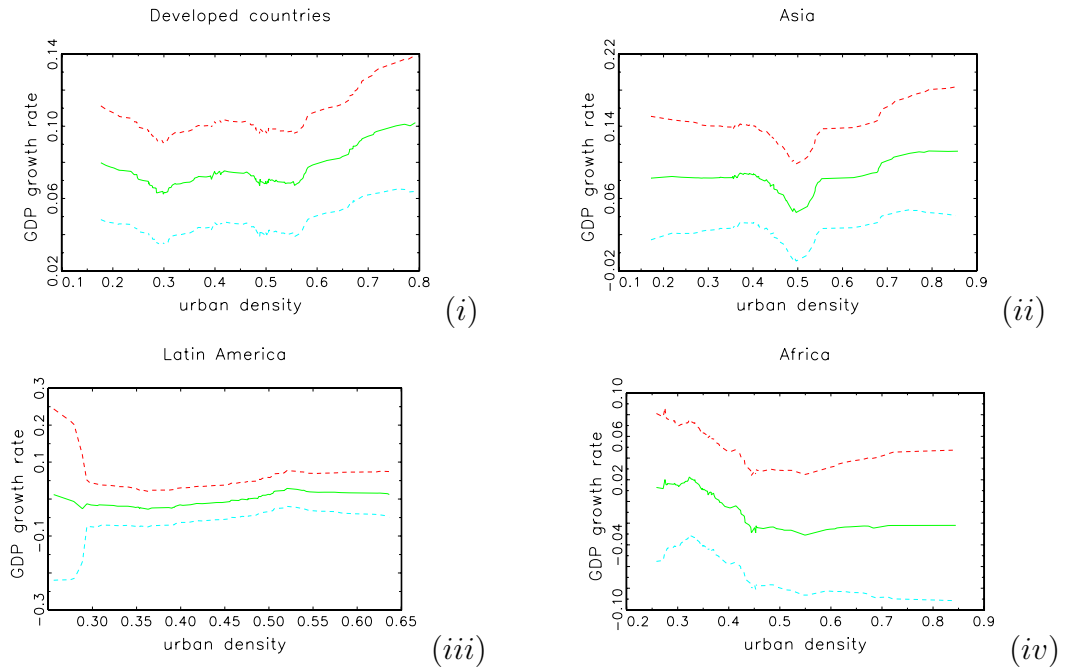


Figure 13: Non-parametric curves with urban density. IV procedure